

ABSTRACT

A system for determining the start of a match of a regular expression includes a special state table that contains start entries and terminal entries, and a set of start state registers for holding offset information. The system further includes a DFA next state table that, given the current state and an input character, returns the next state. A settable indicator is included in the DFA next state table corresponding to each next state table entry which indicates whether to perform a lookup in the special state table. A compiler loads values into the special state table based on the regular expression. A method for determining the start of a match of a regular expression using the special state table, the set of start state registers and the DFA next state table, includes the step of determining from the regular expression each start-of-match start state and each end-of-match terminal state. For each start state, a start state entry is loaded into the special state table. For each terminal state, a terminal state entry is loaded into each special state table. The next state table is used to return the next state from the current state and an input character. When a start state is encountered, the current offset from the beginning of the input character string is loaded into the start state register. When a terminal state is encountered, the terminal state entry is retrieved from the special state table, and the value of the start state register corresponding to the rule number of the terminal entry in the special state table is further retrieved. The value of the start state register which is retrieved indicates the location in the character string where the start-of-match occurred for a particular rule.